Another example: AI methods in Psychology

The orientation of AI research toward the construction of intelligent agents -- known as "knowledge engineering" -- has always coexisted with an orientation toward the explication and understanding of human cognitive behavior viewed as information processing. Indeed the marriage of AI models and methods with the problems and techniques of Cognitive Psychology has been so fruitful that a field with its own name, society, and journal has been born thereof: Cognitive Science.

Since the health research community has long been a supporter of basic research in Cognitive Psychology through the NIMH, it has been appropriate that this branch of AI be supported by SUMEX. The gains thereby have been perceived to be so significant that the Cognitive Science field is itself now considering the establishment of a network-based community, for which SUMEX is one of the leading two models.

The significance of the AI methodology to the modeling of cognitive processes has always been seen as:

precision of expression...computer programming languages are not only ideally suited for expressing the elementary information processes of the model and the postulated data structures, but admit no vagueness or incompleteness.

complexity...the difficulty of managing the modeling process does not go up significantly as the model becomes richer (more complex); thus the methodology does justice to the complexity of human cognitive processes, does not force oversimplifications.

testability...though the models are complex, the computer will generate in detail the remote consequences of the modeling assumptions for particular situations; thus the models are as testable and correctable, in principle, as any in the "hard" sciences.

In recent years, SUMEX-AIM has been one of the most significant forces impelling the forward motion of cognitive science. It has allowed the building of geographically dispersed communities around a single modeling effort; and it has reduced the "cost of entry" to this methodology.

The best example relates to the ACT model of human long-term associative memory, initially constructed by John Anderson. This elegant model has been explored, modified, and tested by a subcommunity of psychologists who gain access to it by the normal simple SUMEX-AIM procedures (bypassing the laborious process, sometimes impossible to achieve, of "bringing it up" at their own sites). As another example, Professor Kintsch and his group at the University of Colorado were able, on the second day of a visit by two Stanford researchers, to begin the process of using the Stanford-SUMEX-developed system, AGE, to model human story comprehension.

What is the GENERAL SIGNIFICANCE of SUMEX-AIM?

As a Research Resource...

SUMEX-AIM is widely viewed as a model national computing resource. Its service has been wide-ranging, in terms of user help and variety of software services provided; reliable; economical on a per-user or per-project basis; and effective in promoting the healthy growth of its research community. It is being studied by communities of scientists in molecular biology (both in the U.S. and Europe) and in cognitive science as a model of how to provide similar service to their sciences; and the term "SUMEX-like facility" was common in planning discussions for the National Center for Computation in Chemistry and for a proposed ARPA national computing resource for ARPA-sponsored DOD projects.

As an experiment in community building...

Lederberg's original vision extended far beyond the "resource" mandate. He said, in an earlier SUMEX renewal proposal,

> "We infer that many fields of scientific inquiry will have to use similar methods of exchange of critical commentary; that the electronic communications of computer programs is a prototype for the maintenance of other knowledge bases essential for the fabric of a complex and demanding society. The computer is at one time the node of a knowledge-sharing network, and the device for verifying the consistency and pertinence of the updates and criticisms that the users remit. Thus we can view our resource as exemplifying a technology that induces a new social organization of scientific effort."

SUMEX-AIM has been remarkably, though not uniquely, successful in pointing to this new direction for scientific integration and cumulation. The collection of computer science research centers on the ARPANET represents another example, but because the goals of SUMEX are more focused, its achievements at community building are more easily defined. The speed with which the relatively new MOLGEN programs are making their way into the relevant scientific community, by means of help from and access to SUMEX, is gratifying evidence of the community building spirit and technique of the resource. That this path cut by SUMEX in the '70s will become the highway of the 80's and '90s is very likely.

As a focus for the development of the inexpensive "intelligent assistant" in medicine and the biosciences...

Artificial Intelligence is the computer science of symbolic representations of knowledge and symbolic inference. There is a certain inevitability to this branch of computer science and its applications, in

particular, to medicine and biosciences. The cost of computers will fall
drastically during the coming two decades. As it does, many more of the
practitioners of the world's professions will be persuaded to turn to
economical automatic information processing for assistance in managing the
increasing complexity of their daily tasks. They will find, in most of
computer science, help only for those of their problems that have a
mathematical or statistical core, or are of a routine data-processing
nature. But such problems will be rare, except in engineering and physical
science. In medicine, biology, management -- indeed in most of the world's
work -- the daily tasks are those requiring symbolic reasoning with
detailed professional knowledge. The computers that will act as
"intelligent assistants" for these professionals must be endowed with such
reasoning capabilities and knowledge. The researchers of the SUMEX-AIM
community currently constitute a large fraction of all the computer
scientists whose work is aimed at this inevitable development.

The day is not far off. There appeared in Business Week, April 14,
1980 an article on INTEL and their plans for the 1980's. INTEL is
presently fourth in integrated circuit sales but is on a much faster growth
curve than its competitors. Therefore its plans should be an important
indicator of the technological environment to be expected in this coming
decade.

INTEL's plans include a "minimainframe" more powerful than any chip
computer so far announced, which includes the ability to be linked in
networks for even higher performance. INTEL is investing about $100
million in software for a full-fledged operating system with capabilities
in language understanding, mechanization of intellectual activity, pattern
recognition etc..

SUMEX-AIM is laying the scientific base so that medicine will be able
to take advantage of these technological opportunities for inexpensive
computer power. Medical diagnostic aids and tools for the medical
scientist that operate in a environment of a network of VAX-like and
$30,000 "professional workstation" computers have the practical possibility
of large-scale and low-cost use because of these anticipated near-term
industrial developments.

As a focus for the methodology that will explicate and disseminate
the "private" -- heuristic -- knowledge of practice...

Knowledge is power, in the profession and in the intelligent agent.
As we proceed to model expertise in medicine and its related sciences, we
find that the power of our programs derives mainly from the knowledge that
we are able to obtain from our collaborating practitioners, not from the
sophistication of the inference processes we observe them using.
Crucially, the knowledge that gives power is not merely the knowledge of
the textbook, the lecture and the journal but the knowledge of "good
practice" -- the experiential knowledge of "good judgment" and "good
guessing", the knowledge of the practitioner's art that is often used in
lieu of facts and rigor. This heuristic knowledge is mostly private, even
in the very public practice of science. It is almost never taught

explicitly; almost never discussed and critiqued among peers; and most often is not even in the moment-by-moment awareness of the practitioner.

Perhaps the the most expansive view of the significance of the work of the SUMEX-AIM community is that a methodology is emerging therefrom for the systematic explication, testing, dissemination, and teaching of the heuristic knowledge of medical practice and scientific performance. Perhaps it is less important that computer programs can be organized to use this knowledge than that the knowledge itself can be organized for the use of the human practitioners of today and tomorrow.

Lederberg's statement from our previous proposal rounds out this larger view:

"Although our substantive efforts are mostly concerned with the 'micro-problems' of scientific or clinical inference, there may be more important treasures in a macro-perspective on the integration of knowledge in medicine. I believe that it is reasonable to expect that the systematization of biomedical knowledge, to which computer AI will make an indispensable contribution, is an important side effect of these investigations in knowledge-engineering; and that this will lead in turn to the recognition of holes in the overall fabric that badly need patching. We have too little theory of the practice of science to offer more than case studies at this time."

5   Progress

This report covers only the resource nucleus; objectives and progress for individual collaborating projects are discussed in their respective reports in Section 9 beginning on page 135. These projects collectively provide much of the scientific basis for SUMEX as a resource and our role in assisting them has been a continuation of that adopted for the first grant term. Collaborating projects are autonomous in their management and provide their own manpower and expertise for the development and dissemination of their AI programs.

5.1   Brief Statement of Prior Goals

The following summarizes SUMEX objectives for the on-going three year grant, begun on August 1, 1978. It will be noted that the high-level goals for this work closely parallel those for the renewal period. These are the continuing basis for our long-term program in biomedical AI research and are resummarized here to comply with the requested NIH form for this proposal. Changes to previous detailed objectives because of explicit guidelines and funding limits in the council award are noted below.

5.1.1   Resource Operations

1) Continue the building of a community of projects applying AI techniques to medical problems including improving mechanisms for inter- and intra- group collaborations and communications.

2) Provide an effective computing resource to support the development and research dissemination of biomedical AI computer programs for a broad range of applications areas.

3) Provide effective and geographically accessible network communication facilities to the SUMEX-AIM community for remote collaborations, scientific communications, and experimentation with developing AI programs.

5.1.2   Training and Education

1) Provide documentation and assistance in interfacing users to resource facilities and programs.

2) Continue to allocate "collaborative linkage" funds to qualifying new and pilot projects to provide for communications and terminal support pending formal approval and funding of their projects. These funds are allocated in cooperation with the AIM Executive Committee reviews of prospective user projects.

3) Continue to <u>support</u> <u>technical</u> <u>workshop</u> <u>activities</u> in collaboration with the Rutgers Computers in Biomedicine resource and individual application projects.

We had proposed support for a "visiting scientist" position to allow prospective qualified SUMEX-AIM project investigators or users to spend a term in close contact with on-going research work. Funding for this position was cut by the NIH review committees.

5.1.3   <u>Core</u> <u>Research</u>

1) Continue to <u>encourage</u> <u>community</u> <u>efforts</u> <u>at</u> <u>organizing</u> <u>and</u> <u>developing</u> <u>AI</u> <u>techniques</u> by supporting projects such as the AI Handbook, special language developments, and other projects community members may propose to contribute.

2) Explore <u>generalizations</u> <u>of</u> <u>AI</u> <u>tools</u> for knowledge acquisition, representation, and utilization.

3) <u>Explore</u> <u>AI</u> <u>software</u> <u>implementation</u> <u>and</u> <u>export</u> <u>mechanisms</u> such as machine-independent languages and special purpose computer systems. This includes the continued development of the MAINSAIL system and the investigation of satellite general purpose machines capable of running existing systems.

Because of guidelines and funding limits in the council-approved award, we removed several goals in the core research work as originally proposed including support for development of a general planning package, a heuristic knowledge acquisition system, and a general explanation system. We were also forced to limit the goals of the MAINSAIL effort to the completion of the language design and to a demonstration of implementations for five target systems. No export efforts for MAINSAIL or work on microprogrammed implementations were possible.

5.2   Summary of Progress: 11/77 - 4/80

1) We have continued to recruit a growing community of user projects
    and collaborators. The initial complement of 5 projects has grown
    to 17 fully authorized projects currently plus a group of 8 pilot
    efforts in various stages of formulation. Several of these projects
    use the AIM computing facility at Rutgers. Many projects are built
    around the communications network facilities we have assembled,
    bringing together medical and computer science collaborators from
    remote institutions and making their research programs available to
    still other remote users.

2) SUMEX user projects have made good progress in developing and
    disseminating effective consultative computer programs for
    biomedical research. These performance programs provide expertise
    in analytical biochemical analyses and syntheses, medical diagnoses,
    and various kinds of cognitive and affective psychological modeling.
    We have worked hard to meet their needs and are grateful for their
    expressed appreciation. [see Section 9 beginning on page 135].

3) A first version of the AGE system has been completed. It uses the
    "blackboard model" control structure for coordinating multiple
    expert sources of knowledge for the solution of problems. The UNITS
    package [9] for a "frame-oriented" representation of knowledge is
    now being incorporated. AGE provides a general structure and an
    interactive facility for implementing knowledge-based systems. A
    workshop to introduce AGE to the AIM community was held at Stanford
    in February 1980. [see Section 9.1.1 on page 137].

4) We have completed the initial phases of a systematic effort to
    document AI concepts and techniques through the AI Handbook Project.
    It comprises a compendium of short articles about the projects,
    ideas, problems, and techniques that make up the field of AI. The
    first two volumes covering heuristic search, knowledge
    representation, natural language and speech understanding, AI
    languages, various applications domains, and automatic programming
    were completed in August 1979 and publication plans are in progress.
    All completed sections have been published as Stanford Computer
    Science Department technical reports. Work on a third volume is
    progressing well. [see Section 9.1.2 on page 145 and Appendix G
    on page 392]

5) We successfully completed the design and a demonstration of the
    MAINSAIL language system as a tool for software portability. A
    common compiler, code generators, and runtime support for TENEX,
    TOPS-10, TOPS-20, RT-11, and RSX-11 have been developed as part of
    this demonstration system and numerous applications programs written
    by collaborating research groups. Further work past this
    demonstration phase will be done independently of SUMEX through a
    private company, XIDAK, formed to continue the development,
    dissemination, and maintenance of MAINSAIL. Work is under way to
    develop MAINSAIL for the VAX and a number of other target machines.
    [see Appendix H on page 398].

6) We have continued refinement of the SUMEX facility hardware and
    software systems. We have worked to enhance throughput, to better
    control the allocation of resources among communities, to increase
    efficiency, to enhance human interfaces, to improve documentation,
    and to extend the range of software facilities available to user
    projects. We also completed installation and evaluation of a
    connection to TELENET as an alternate source of communications
    services for our community.

7) We completed planning and implementation of a satellite machine that
    supports more operational demonstrations of mature AI programs and
    helps alleviate system congestion for on-going program development.
    This acquisition of a DEC 2020 system was reviewed and approved by
    an ad hoc study section. We have installed the machine and are
    actively working on its integration into KI-10 facility by means of
    a local Ethernet [10]. Using an interim connection, it has been
    used extensively for workshops and program demonstrations.

8) We have smoothly completed the management transition. On July 1,
    1978, Prof. Edward Feigenbaum assumed the role of SUMEX Principal
    Investigator following Prof. Joshua Lederberg's installation as
    president of The Rockefeller University. Prof. Lederberg continues
    to maintain close ties with SUMEX activities as chairman of the
    SUMEX-AIM Executive Committee. Close coordination of project
    activities with medical research is provided by Dr. E. H.
    Shortliffe, co-Principal Investigator of SUMEX. Dr. Shortliffe is
    Assistant Professor of General Internal Medicine and one of the key
    developers of the MYCIN system. Effective August 1, 1980, SUMEX
    will become part of the Department of Medicine where it will be
    centered in the largest clinical department of the Stanford Medical
    School. Previously, SUMEX had been in the Department of Genetics
    with Prof. Stanley Cohen, Dr. Lederberg's successor as chairman,
    assisting in project medical coordination.

5.3   Detailed Progress Highlights

     The following material highlights in more detail SUMEX-AIM resource
activities since the last review in the context of the resource staff and
the resource management.

5.3.1   Resource Operations

     Our core facility, initially installed in March 1974, is built around
a Digital Equipment Corporation (DEC) KI-10 computer and the TENEX
operating system. This facility has provided a superb base for the AI
mission of SUMEX-AIM in terms of its interactive computing environment, its
AI program development tools, and its network and interpersonal
communication media. Biomedical scientists have found SUMEX easy to use in
exploring applications of developing artificial intelligence programs for
their own work and in stimulating more effective scientific exchanges with
colleagues across the country.

     These tools also give us access to a large computer science research
community, including active artificial intelligence and system development
research groups. Coupled through effective network facilities, these
groups greatly enhance the SUMEX-AIM community environment through broader
scientific interchange and software sharing.

     Following are highlights for recent developments in various aspects
of the facility. Detailed information about SUMEX loading can be found in
Appendix B on page 355. Plots are given there for overall resource usage,
diurnal loading, community/project usage, and network traffic.

5.3.1.1   System Hardware

   1) Implemented a number of strategic facility augmentations over the
       years in response to growing community needs to increase system
       capacity and improve performance for interactive expert systems.
       These include: (3/74) - install KI-10 with 192K words of memory;
       (11/74) - add 64K words of memory; (5/76) - add second KI-10; (8/77)
       - add 256K words of memory and double on-line file space (see Figure
       1 for a current configuration diagram).

   2) Acquired a software-compatible satellite DEC 2020 computer as a
       dedicatable resource for improved interactive response for
       experimental testing of AI programs. This relatively inexpensive
       machine ($175,000) includes a KS-10 processor approximately half the
       speed of a KI-10, 512K words of memory, 1 disk and 1 tape drive, 16
       terminal lines, and software license (see Figure 2 for a
       configuration diagram}). It runs TOPS-20 and is for the most part
       software-compatible with the KI-TENEX system. The 2020 was
       installed without problem in August 1979 and we have supported many
       program demonstrations on it for the DENDRAL, ONCOCIN, AGE, SECS,

INTERNIST, and MOLGEN projects. Major conferences for which the
2020 has been used include the Sixth International Joint Conference
on AI from Tokyo, Japan in August 1979 and, most recently, the
American College of Physicians meeting in New Orleans in April 1980.

3) Began implementation of a local Ethernet [10] as the basis for
    integrating the KI-10 facility with the 2020 and future planned
    hardware. Based on Xerox-developed protocols, this system will
    connect SUMEX resources through a 3.3 Mbit/sec network to allow
    uniform terminal access, file transfers, peripheral equipment
    sharing, and remote resource access through gateways. Figure 3 on
    page 38 shows current configuration plans for the SUMEX network.
    The KI-10's are fully operational on the Ethernet through an interim
    I/O bus PDP-11 interface. This uses a Xerox-designed PDP-11
    interface board and an adaptation of their higher level software.
    The 2020 is connected electrically through its UNIBUS adapter. We
    are working to complete the 2020 connection software and to design a
    direct memory interface for the KI-10's to achieve higher
    performance and efficiency. [see Appendix C on page 374 for
    details].

4) We have designed and implemented communications control hardware to
    allow sensing of carrier drop on dial-up lines so that attached jobs
    can be detached to prevent users from inadvertently connecting to
    hanging jobs. We also implemented a software-controlled switch to
    allow more efficient use of available terminal scanner ports on the
    system. Hardwired and leased line connections no longer tie up
    scanner ports when not in use.

5) We have supported community hardware communication needs by
    installing and maintaining local terminals and connections;
    assisting in the acquisition and installation of terminals at remote
    user sites; assisting with dedicated links to remote user sites
    (e.g., UC Santa Cruz and UC San Francisco); and assisting with
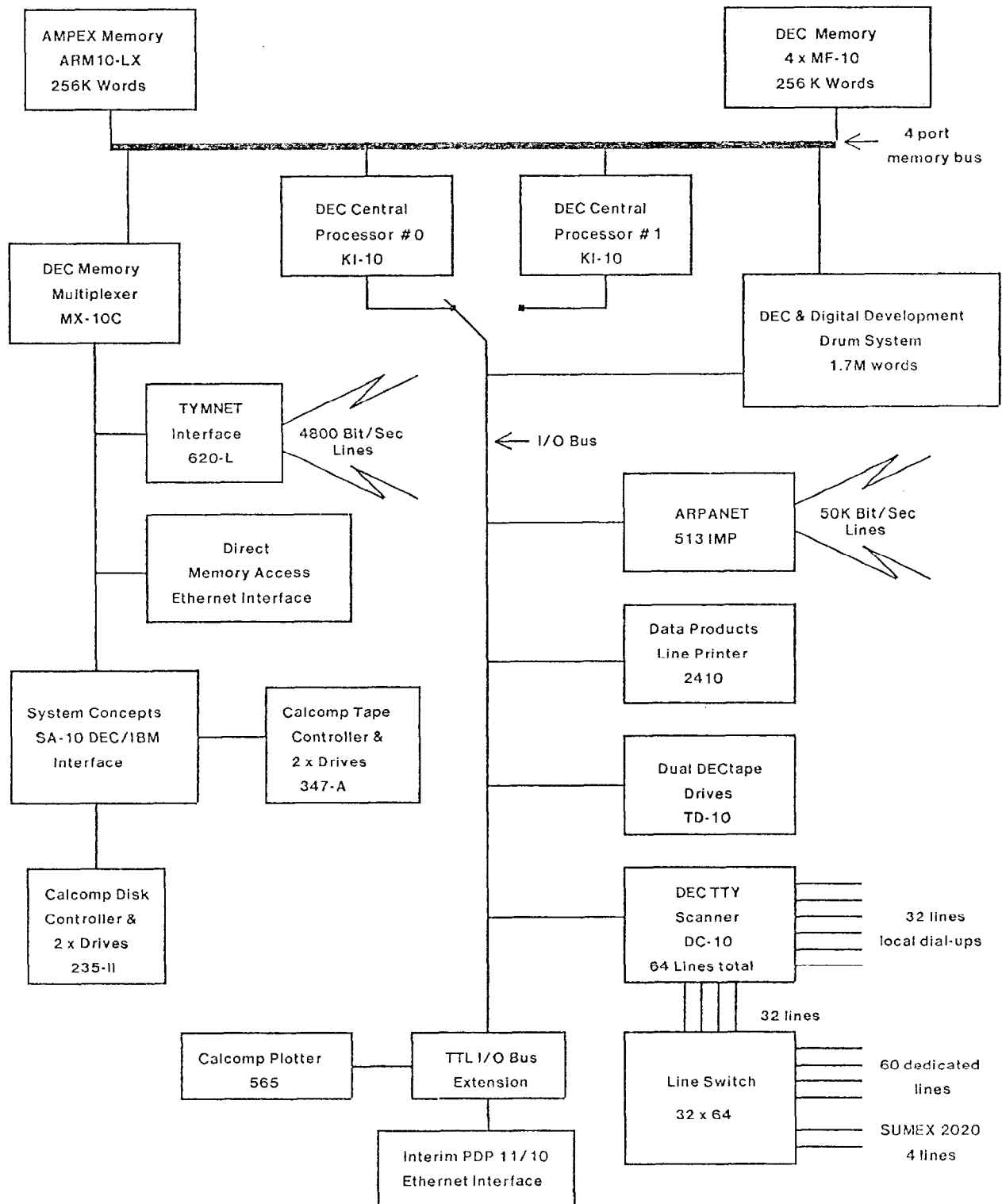   equipment installation for AI program demonstrations.

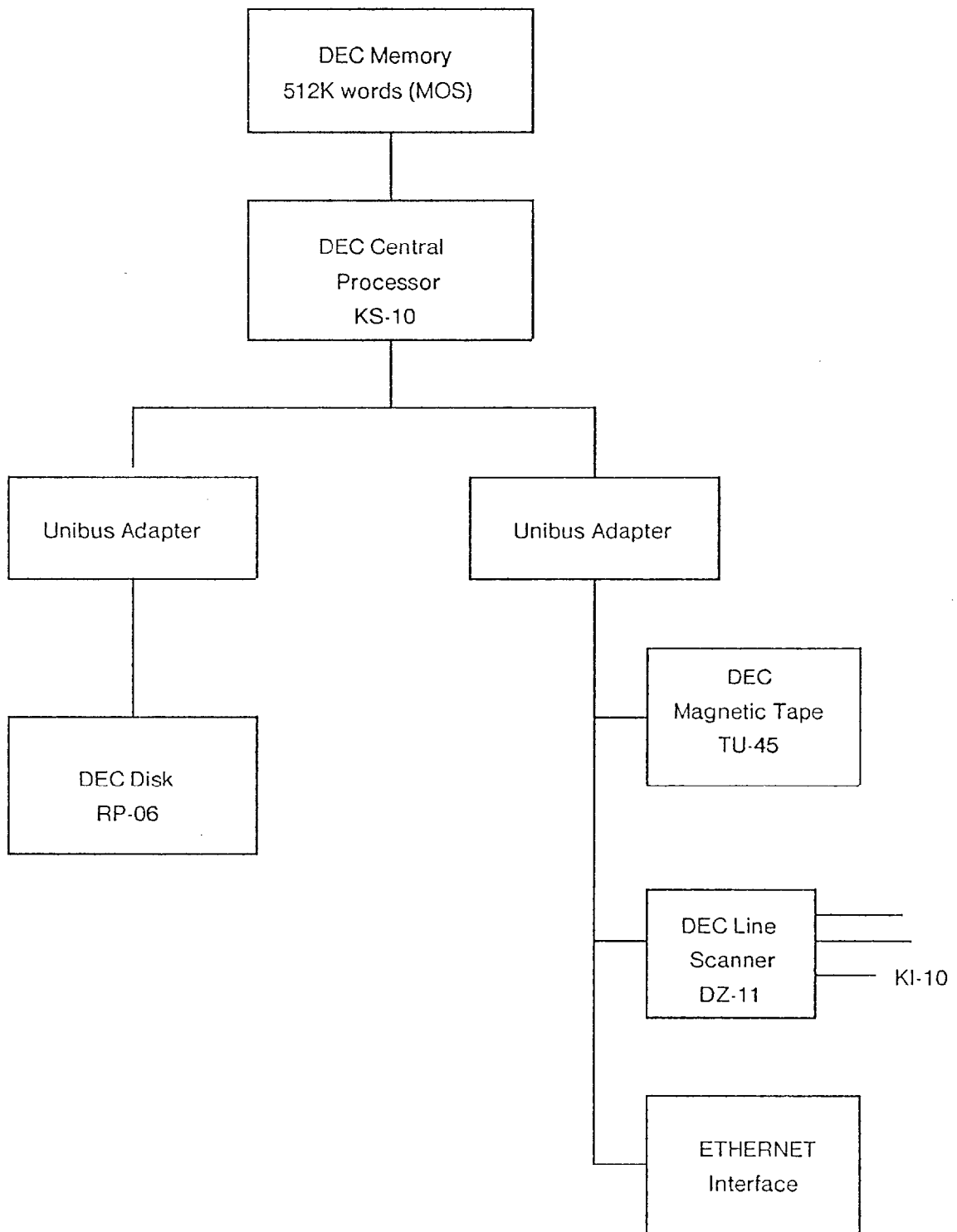Figure 1.   Current SUMEX-AIM KI-10 Computer Configuration

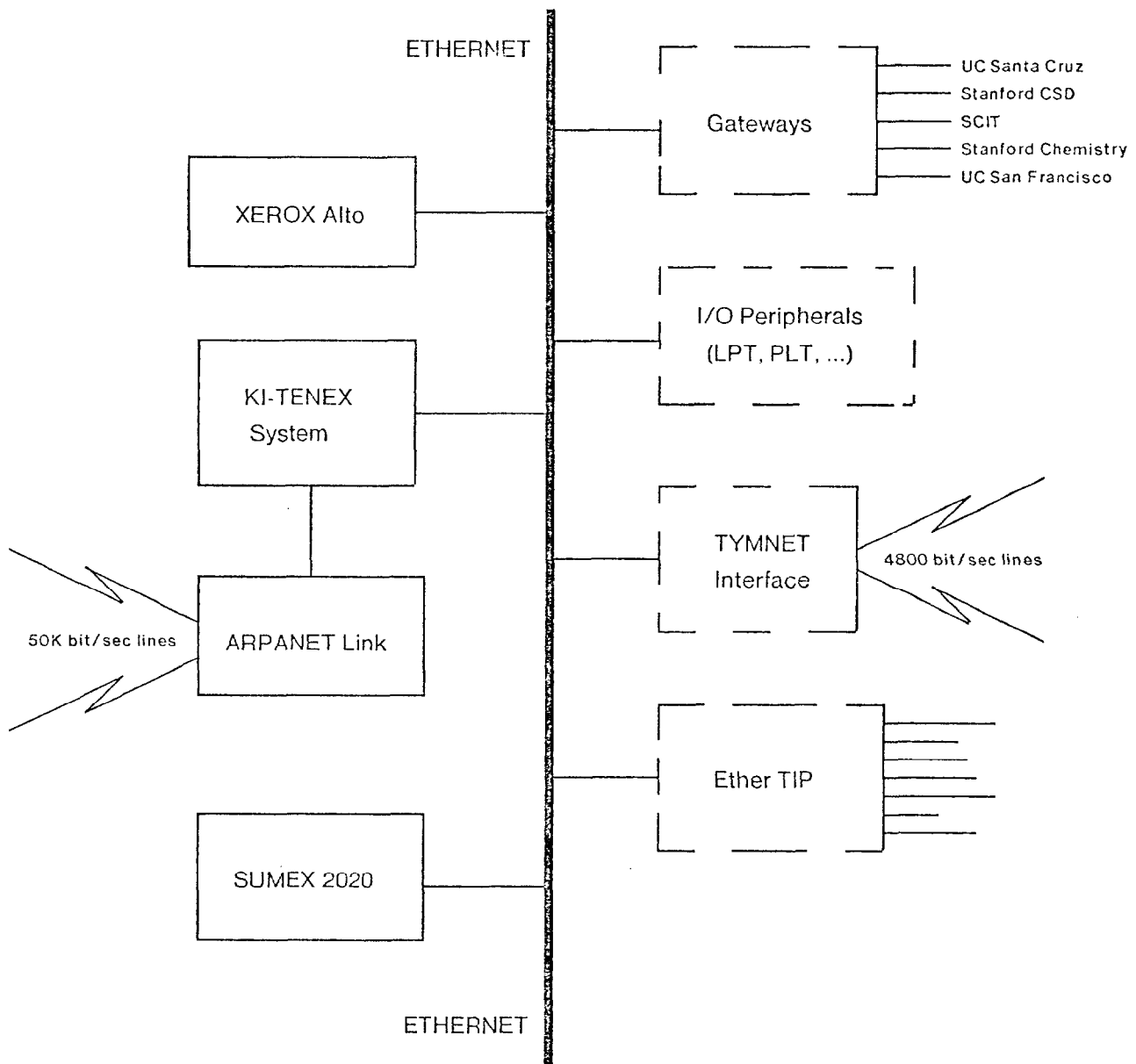Figure 2.   Current SUMEX-AIM 2020 Computer Configuration

ETHERNET

Gateways
— UC Santa Cruz
— Stanford CSD
— SCIT
— Stanford Chemistry
— UC San Francisco

XEROX Alto

I/O Peripherals
(LPT, PLT, ...)

KI-TENEX
System

TYMNET
Interface

4800 bit/sec lines

50K bit/sec lines

ARPANET Link

Ether TIP

SUMEX 2020

ETHERNET

Figure 3.   Intermachine Connections via ETHERNET

5.3.1.2   System Software

In parallel with the choice of DEC PDP-10 hardware for the SUMEX-AIM facility, we selected the TENEX operating system developed by Bolt, Beranek, and Newman (BBN) as the most effective for our medical AI applications work. Together with the hardware, TENEX has provided a superb environment in which to pursue community biomedical AI applications work. Following are highlights of recent system software developments:

Monitor

1) we have made significant contributions to the KI-TENEX monitor that are now in use at other sites. These include efficiency improvements in the management of user page tables, implementation of a memory-shared TYMNET interface including outbound circuit facilities, design and implementation of the dual processor TENEX system, implementation of a page migration system to assure effective use of fixed-head swapping storage, and improvements in system routines for locating and recognizing file names.

2) developed overload control facilities that effectively limit the number of active processes on the system to those that can be supported with reasonable response time. These provide for "background" jobs, "demo priority" jobs, and mechanisms to temporarily suspend user jobs that have not cooperated with requests to reduce the system load. Active process slots are allocated on the basis of a priori resource percentages that communities and projects are entitled to.

3) implement monitor communication controls for the experimental TELENET network connection. These included special "Xon/Xoff" facilities to allow transmission of packets into the network at 1200 baud irrespective of terminal speed so that network transmission delays could be minimized. Network "backpressure" commands prevented overruns for slower terminals. [see Appendix D on page 376 for details].

4) implement monitor service routines for the "carrier detect" control and line switching hardware.

5) examined KI-TENEX page faulting behavior to measure the utility of block transferring pages in anticipation of faults. Data for a wide range of programs indicate that TENEX already does a good job of keeping needed pages in memory, limited by the amount of physical memory available. We propose to add another 256K of core memory to the system to reduce swapping overhead.

6) integrate the Ethernet and PUP monitor service routines adapted from Xerox PARC [10, 13]. This required redesigning the hardware interface code for our interim PDP-11 I/O bus interface (KI-10) and the 2020 18-bit UNIBUS adapter, changing executive "XCT" codes to conform to differences in hardware function between the Xerox microcoded PDP-10 and our KI-10's, and implementing needed

additional system calls (JSYS's). The KI-10 is fully working on our
Ethernet. The extensive TOPS-20 monitor changes for the 2020 are
still in progress.

7)   adapt the TOPS-20 monitor from the Stanford DEC 2060 systems to the
     SUMEX 2020. We have made minimal changes to the monitor code except
     to accommodate the Ethernet interface and to provide needed controls
     for priority program demonstration and testing.

8)   make numerous monitor bug repairs to provide for more reliable
     system operation and file integrity. Obvious bugs were removed long
     ago so those remaining are elusive and occur infrequently. We have
     found and fixed bugs in the management of multi-fork structures, the
     ARPANET control programs, the file page backup routines, the
     manipulation of special monitor pages mapped through the user page
     table, and the concatenation of drum I/O requests for latency
     reduction.

Utility Features

     We have made a significant number of utility improvements to the
monitor to add new features, improve compatibility with TENEX 1.34 and
TOPS-20, or improve operational effectiveness. A brief list includes:

1)   Printer device and spooler that manages a print queue for Prof.
     Wipke's group at UC Santa Cruz. This device allows interspersing
     use of the UCSC link as a terminal line and as a printer device.

2)   Password error monitoring to log out jobs causing a high number of
     failures and to report the source and target directories to the
     operator. This is designed to catch occasional attempts at
     unauthorized entry into the system, generally from remote network
     connections.

3)   Improved GTJFN features to partially recognize ambiguous file names
     up to the point of ambiguity and to recognize parts of the TOPS-20
     name syntax for compatibility.

4)   Upgrade routines and JSYS's to conform with TENEX 1.34 to provide
     desirable new features (selective expunge, group connect, improved
     file system physical format, and expanded directory hash table) and
     to retain compatibility with evolving ARPANET protocols.

5)   Checksum monitor code as loaded to detect I/O device errors or
     memory problems.

6)   Make the console teletype of the second processor available for use
     and improve operational procedures for taking crash dumps and
     reloading the system.

System Executive

One of the most important system programs is the EXECutive which is the basic user interface to manipulate files, directories, and devices; control job and terminal parameter settings; observe job and system status; and execute public and private programs. The SUMEX EXEC is quite well developed at this stage but we have made several recent improvements:

1) Implementation of LOGIN.CMD and COMAND.CMD files which are processed at login and upon starting any new EXEC. These files allow the user to give any available EXEC command automatically to set default parameters, print status information, etc.

2) Enhancement of the functions and improvement of the human interaction of the file archive/retrieval system. Users can now specify a list of files to be retrieved, edit their archive directories to remove old entries or collect groups of entries, annotate entries to better document contents, and interactively step forward and backward when searching for an entry.

3) Implementation of general wild card facilities for the COPY and RENAME commands. This allows users to copy/rename groups of files to new files with names derived by reorganizing selected substrings from the originals thereby reducing the manual typing required.

4) Implement the selective expunge command from TENEX 1.34 so that temporary files (e.g., MESSAGE.COPY) can be retained while expunging unneeded deleted files.

5) Improvement of the scheduling control information provided to users for planning their work around overloaded system conditions.

6) Implement demo controls for the 2020 EXEC to preserve its capacity during scheduled sessions for AI program tests or demonstrations.

System Utilities and Operations

We have made numerous improvements and bug fixes to the system utility and operations programs needed to assist smooth management of the system and to provide new facilities for users. A brief list of the most significant tasks includes:

1) Spooler improvements - allow users to retract requests to list files and implement a special spooler for printing files remotely at UC Santa Cruz for Prof. Wipke's group. This spooler communicates over a line also used for terminals and uses a specially designed protocol to coordinate line usage.

2) SYSJOB controls - several of the system utilities for TELNET connections, mail forwarding, statistics collection, TYMNET downtime msg updating, etc. were relocated to a separate system job to facilitate better resource allocation controls and to reduce

competition with other critical system functions (disk page backup
and network control programs).

3) <u>Overload controls</u> - implement the user-level demo priority and
uncooperative job controls for overloaded system conditions based on
the monitor control functions described earlier.

4) <u>File archive/retrieval</u> - improvements to BSYS incorporating user
status information on retrieval processing and the latest BBN system
for file restoration automation.

5) <u>File system verification</u> - improvements to the CHECKDSK program for
detecting file system integrity problems after a crash to allow
better notification to users of the names of files that might have
been lost or damaged.

6) <u>System and crash analysis</u> - improvements to the program developed to
assist in sorting through the complex interlinked monitor tables
when unraveling a core dump to analyze the cause of a crash. Also
develop several display programs to observe the dynamic operation of
individual job structures or network connections.

7) <u>Ethernet/PUP service</u> - import and adapt to the SUMEX system the
Xerox user-level service programs for file transfer, terminal
connections, mail forwarding, gateway routing, etc.

8) <u>2020 conversions</u> - on-going conversion of useful KI-10 programs to
run in the TOPS-20 environment.

9) <u>TENEX/TOPS-20 compatibility package</u> - we have made substantial
extensions to a compatibility package, PA-2040, that was originally
written at USC-ISI. This package now emulates many of the TOPS-20
unique JSYS's. We have added the monitor mode instruction emulation
software written initially for the SUMEX GTJFN development so that
unique TOPS-20 monitor JSYS code can be run directly from user
space. This allows JSYS's without TENEX equivalents to be emulated
directly. There are still TOPS-20 JSYS definition changes that
cannot be handled by means of a compatibility package.

User Subsystems

We have continued to assemble (develop where necessary) and maintain
a broad range of user support software. These include such tools as
language systems, statistics packages, DEC-supplied programs, improvements
to the TOPS-10 emulator, text editors, text search programs, file space
management programs, graphics support, a batch program execution monitor,
text formatting and justification assistance, magnetic tape conversion
aids, and user information/help assistance programs.

1) <u>new installations or versions of subsystems</u> essential to users have
been brought up with varying requirements for local adaptation to
run on the SUMEX KI-10's. New or updated subsystems include MLAB

and OMNIGRAPH from NIH; FORTRAN, CCL, COBOL, BACKUP, MACRO, LINK10, GLOB, and a new set of utility routines used by many of the DEC CUSP's from DEC; INTERLISP from Xerox PARC; ESSEX-BCPL from the University of Essex in England; PASCAL and SAIL from Rutgers University (C. Hedrick); PUB (a text formatting program) from IMSSS (M. Hinckley) and SUMEX; MSG (a mail reading program) from BBN (J. Vittal); and TEX (a text publication system) from Stanford (D. Knuth).

2) upgrade the crt display package in the TV text editor to support many additional terminals. TV now handles Teleray-1061, Heath H-19, and a locally modified version of the Hazeltine 1500. Support will soon be available for the NIH Delta Data 5200, Infoton 400, and Visual 200. We are also incorporating enhancements made recently by C. Hedrick at Rutgers to allow improved search and text relocation facilities.

3) import and support the EMACS text editing system from MIT. Substantial effort has gone into developing macro packages that improve the human engineering features of EMACS and providing introductory documentation for new users. This has been closely coordinated with similar efforts at SRI and MIT. A community of EMACS users is now developing at SUMEX.

4) add features to allow attaching batch jobs that have an initial interactive phase that has to be run from a user terminal but which can then be turned over to batch operation for background or deferred running. Also improve batch efficiency and help facilities.

5) add facilities to the spelling corrector to replace misspelled words with phrases, remember the names of subdictionaries loaded, and override misspellings to do simple translations.

Communications Subsystems

Of key importance for our community effort is a set of tools for inter-user communications. We have built up a group of programs to facilitate many aspects of communications including interpersonal electronic mail, a "bulletin board" system for various special interest groups to bridge the gap between private mail and formal system documents, and tools for terminal connections and file transfers between SUMEX and various external hosts. Recent developments include:

1) TTYFTP - A system for file transfers usable over any circuit that appears as a terminal line to the operating system (hardline, dial-up, TYMNET, etc.) and incorporating appropriate control protocols and error checking. The design is derived from the DIALNET protocols developed at the Stanford AI Laboratory with extensions to allow both user and server modules to run as user processes without operating system changes. TTYFTP is written in MAINSAIL and is implemented for TENEX, TOPS-20, RT-11, and RSX-11M.

2) Bulletin Board - BBD has been extended to allow remote posting of bulletins via communication network and has improved efficiency.

3) VTTY - we have combined outbound (TELNET) terminal access protocols for TYMNET, SCIT (Stanford IBM facility), SUMEX 2020, and pseudoteletypes in a single virtual terminal program. VTTY provides typescript services to record sessions.

4) Electronic mail - improve the mail facilities for guests and allow reediting of all message fields (i.e., addressees, subject, and body) in SNDMSG. Also import the more efficient protocols for network mail developed by K. Harrenstien at MIT.

Software Sharing

At SUMEX-AIM we are committed to importing rather than reinventing software where possible. As noted above, a number of the packages we have brought up are from outside groups. Many avenues exist for sharing between the system staff, various user projects, other facilities, and vendors. The availability of fast and convenient communication facilities coupling communities of computer facilities has made possible effective intergroup cooperation and decentralized maintenance of software packages. The TENEX sites on the ARPANET have been a good model for this kind of exchange based on a functional division of labor and expertise. The other major advantage is that as a by-product of the constant communication about particular software, personal connections between staff members of the various sites develop. These connections serve to pass general information about software tools and to encourage the exchange of ideas among the sites.

1) We continue to import significant amounts of system software from other ARPANET sites, reciprocating with our own local developments. Interactions have included mutual backup support, experience with various hardware configurations, experience with new types of computers and operating systems, designs for local networks, operating system enhancements, utility or language software, and user project collaborations.

2) We have assisted groups that have interacted with SUMEX user projects get access to software available in our community. For example, Prof. Dreiding's group in Switzerland became interested in some of the system software available here after attending the DENDRAL CONGEN workshops (see Section 9.1.3 on page 149). We have provided him with the non-licensed programs requested. We are working on a similar arrangement for a group interested in the MOLGEN program.

User Assistance and Documentation

The SUMEX resource exists to facilitate biomedical artificial intelligence applications from program development through testing in the target research communities. This user orientation on the part of the

facility and staff has been a unique feature of our resource and is
responsible in large part for our success in community building.

1) We have <u>tailored resource policies to aid users</u> whenever possible
within our research mandate and available facilities. Our approach
to system scheduling, overload control, file space management, etc.
all attempt to give users the greatest latitude possible to pursue
their research goals consistent with fairly meeting our
responsibilities in administering SUMEX as a national resource.

2) The <u>resource staff has spent significant effort in assisting users</u>
gain access to the system and use it effectively. We respond
promptly to questions by telephone, terminal link, or electronic
mail. We also exercise great care in managing system file integrity
and assisting users in recovering files lost through user error or
system malfunction.

3) We have <u>worked hard to assist projects</u> achieve their goals in
setting up an appropriate computing environment on the system
including directory groups, collaborator and guest facilities, file
space allocations, and special software subsystems.

4) We have <u>solicited and acted upon user recommendations for system
development goals</u>. A "gripe" system is available to users for
general comments as well as electronic mail to individual staff
members responsible for particular aspects of the system.

5) We have spent substantial effort to <u>develop, maintain, and
facilitate access to documentation</u> so as to accurately reflect
available software. The HELP and Bulletin Board subsystems have
been important in this effort. As subsystems are updated, we
generally publish a bulletin or small document describing the
changes. We have worked to review the existing documentation
system, reorganize it for easier access and maintenance, create
command and documentation summaries where appropriate for new users,
and update on-line and hardcopy documents for compatibility with the
programs now running. We have collected useful comparisons and
difference summaries between the KI-TENEX and 2020 systems to assist
users in moving easily between them. Maintenance of accurate and
useful documentation is a continuing task.

5.3.1.3   Network Communication Facilities

A highly important aspect of the SUMEX system is effective
communication with remote users. In addition to the economic arguments for
terminal access, networking offers other advantages for shared computing.
These include improved inter-user communications, more effective software
sharing. uniform user access to multiple machines and special purpose
resources, convenient file transfers, more effective backup, and co-
processing between remote machines. These issues become even more
important with the emerging computing technology that will make increasing
decentralization possible. Networks will be crucial for maintaining the
collaborative scientific and software contacts built up. A detailed
description of our network connections can be found in Appendix D on page
376. Recent milestones include:

1) We continue our connection to TYMNET as the primary means for access
    to SUMEX-AIM from research groups around the country and abroad.
    There has been no significant change in user service or network
    performance. Very limited facilities for file transfer exist and no
    improvements appear to be forthcoming soon. Services continue to be
    purchased through the NLM contract and we have elected "dedicated
    port" pricing as the most cost effective. We continue to have
    serious difficulties getting needed service from TYMNET for
    debugging network problems. See Figure 18 on page 379 for a recent
    list of TYMNET access nodes.

2) We continue our advantageous connection to the Department of
    Defense's ARPANET, now managed by the Defense Communications Agency
    (DCA). Terminal access restrictions are in force so that only users
    affiliated with DoD-supported contractors may use TELNET facilities.
    ARPANET is the primary link between SUMEX and other machine resource
    such as Rutgers-AIM. Current ARPANET geographical and logical maps
    are shown in Figure 19 and Figure 20 on page 380.

3) We implemented an experimental connection to TELENET via a TP-2200
    interface with 12 asynchronous lines to SUMEX and one 4800 baud line
    connecting to the network backbone. In spite of potential economic
    advantages, this experiment was unsuccessful. Users complained of
    poor node reliability, intolerable delays in response, uneven flow
    of terminal output, and poor operational management of the network.
    Similar problems existed from the system standpoint. Other half-
    duplex users (e.g., the NLM MEDLINE system) have reported more
    successful connections. Because of funding limitations, we had to
    abandon our TELENET link for the time being. See Figure 21 on page
    382 for a recent list of TELENET access nodes.

5.3.1.4   Resource Management

        Early in the design of the SUMEX-AIM resource, a rather elaborate
management plan was worked out with the Biotechnology Resources Program at
NIH to assure fair administration of the resource for both Stanford and
national users and to provide a framework for recruitment and development
of a scientifically meritorious community of application projects. This
structure is described in some detail in Appendix E on page 383. It has
continued to function effectively as summarized below.

   1)   The AIM Executive Committee meets regularly by teleconference to
        advise on new project access applications, discuss resource
        management policies, plan workshop activities, and conduct other
        community business. The Advisory Group meets together at the annual
        AIM workshop to discuss general resource business and individual
        members are contacted much more frequently to review project
        applications. (See Appendix I on page 399 for a current listing of
        AIM committee membership).

   2)   Effective July 1, 1978, Prof. Edward Feigenbaum, Chairman of the
        Stanford Department of Computer Science became SUMEX principal
        investigator after Prof. Joshua Lederberg assumed the presidency of
        The Rockefeller University. This transition took place smoothly
        because of Prof. Feigenbaum's role as co-Principal Investigator of
        SUMEX from its start and his long standing collaboration with Prof.
        Lederberg. Close scientific and administrative ties are maintained
        with the Stanford medical community through Prof. Edward H.
        Shortliffe, who is one of the key designers of MYCIN and co-
        Principal Investigator of SUMEX. The project will become
        administratively part of. the Stanford Department of Medicine,
        effective August 1980. As part of the largest clinical medicine
        department at Stanford, SUMEX will have increased visibility and
        opportunity to broaden its local scientific collaborations.

   3)   We have actively recruited new application projects and disseminated
        information about the resource. The number of formal projects in
        the SUMEX-AIM community has nearly quadrupled since the start of the
        project (see Figure 6 on page 331). Here, for example, are just
        some recent efforts to broaden outside awareness of work in the AIM
        community and to encourage new projects: the CONGEN workshop at
        Stanford (1978); the AGE workshop at Stanford (1980); an AI session
        at the Fourth Illinois Conference on Medical Information Systems
        (1979); INTERNIST and MYCIN participation in a course on AI
        computing at NIH (1979); an AI session at the Association for
        Information Science meeting (1979); an AI session at the Sixth
        International Joint Conference on AI (1979); an extensive lecturing
        tour among Japanese university, government, and industrial research
        groups; and MYCIN and INTERNIST program demonstrations at the
        American College of Physicians meetings (1979 and 1980).

   4)   With the advice of the Executive Committee, we have awarded pilot
        project status to promising new application projects and
        investigators and where appropriate, offered guidance for the more

effective formulation of research plans and for the establishment of research collaborations between biomedical and computer science investigators.

5) We have <u>welcomed</u> <u>a</u> <u>number</u> <u>of</u> <u>visiting</u> <u>investigators</u> <u>at</u> <u>Stanford</u> who were able to pay their own expenses, so they could see first hand how AI applications programs are formulated and get acquainted with the computing tools available. Funds for such visiting scientists were deleted from our previous grant award.

6) We have <u>allocated</u> <u>limited</u> <u>"collaborative</u> <u>linkage"</u> <u>funds</u> as an aid to new projects or collaborators with existing projects to support terminals, communications costs, and other justified expenses to establish effective links to the SUMEX-AIM resource. Executive Committee advice is used to guide allocation of these funds.

7) We have <u>carefully</u> <u>reviewed</u> <u>on-going</u> <u>projects</u> with our management committees to maintain a high scientific quality and relevance to our biomedical AI goals and to maximize the resources available for newly developing applications projects. Several pilot projects have been terminated as a result and more productive collaborative ties established for others.

8) We have <u>continued</u> <u>to</u> <u>provide</u> <u>active</u> <u>support</u> <u>for</u> <u>the</u> <u>AIM</u> <u>workshops</u>. The last one was held in May 1979. It was organized by MIT-Tufts and Rutgers and was devoted to clinical diagnosis programs. We also have <u>supported</u> <u>individual</u> <u>project</u> <u>workshops</u> such as those held for CONGEN and AGE. The next AIM workshop will be held at Stanford in August 1980 together with several tutorial sessions on AI for physicians. Prof. Shortliffe is the program chairman for this workshop.

9) We have <u>continued</u> <u>our</u> <u>policy</u> <u>of</u> <u>no</u> <u>fee-for-service</u> for projects using the SUMEX resource. This policy has effectively eliminated the serious administrative barriers that would have blocked our research goals of broader scientific collaborations and interchange on a national scale within the selected AIM community. In turn we have responded to the correspondingly greater responsibilities for careful selection of community projects of the highest scientific merit.

5.3.2   Core Research

Since the last report we have supported several core research
activities aimed at developing information resources, basic AI research,
and tools of general interest to the SUMEX-AIM community. Specific areas
of current effort include:

1) The AI Handbook, under Prof. Feigenbaum and Mr. Avron Barr: a
    compendium of knowledge about the field of artificial intelligence
    being compiled by students and investigators at several research
    facilities across the nation. The handbook is broad in scope,
    covering all of the important ideas, techniques, and systems
    developed during 20 years of research in AI in a series of articles.
    Each is about four pages long and is a description written for non-
    AI specialists and students of AI. The first two volumes covering
    heuristic search, knowledge representation, natural language and
    speech understanding, AI languages, various applications domains,
    and automatic programming are complete. All completed sections are
    published as Stanford Computer Science Department technical reports.
    Work on a third volume is progressing well. [see Section 9.1.2
    on page 145 for a more detailed report and Appendix G on page 392
    for an outline of the handbook contents)

2) The AGE project: an attempt to isolate inference, control, and
    representation techniques from previously developed knowledge-based
    programs; reprogram them for domain independence; write a rule-based
    interface that will help a user understand what the package offers
    and how to use the modules; and make the package available to other
    members of the AIM community. A first version of the AGE system has
    been completed. It uses the "blackboard model" control structure
    for coordinating multiple expert sources of knowledge for the
    solution of problems. The UNITS package [9] for a "frame-oriented"
    representation of knowledge is now being incorporated. AGE provides
    a general structure and an interactive facility for implementing
    knowledge-based systems. A workshop to introduce AGE to the AIM
    community was held at Stanford in February 1980. [see Section
    9.1.1 on page 137 for a more detailed report].

3) The MAINSAIL project: an effort to design and demonstrate a machine-
    independent, ALGOL-like language system to facilitate software
    transportability between different machine/operating system
    environments. We successfully completed the design and a
    demonstration of the MAINSAIL language system as a tool for software
    portability [14, 15]. A common compiler, code generators, and
    runtime support for TENEX, TOPS-10, TOPS-20, RT-11, and RSX-11 have
    been developed as part of this demonstration system and numerous
    applications programs written by collaborating research groups.
    Further work past this demonstration phase will be done
    independently of SUMEX through a private company, XIDAK, formed to
    continue the development, dissemination, and maintenance of
    MAINSAIL. Work is under way to develop MAINSAIL for the VAX and a
    number of other target machines. [See Appendix H on page 398 for a
    more detailed summary of the final phases of this project].

It should be noted that SUMEX provides only partial support for the
AI Handbook and the AGE projects with complementary support coming from an
ARPA contract to the Heuristic Programming Project. Other portions of our
original proposal for core research in knowledge acquisition, planning, and
generalized explanation systems have not been supported for lack of
resources following council reduction of this section of our budget.